Using Deep NLP to Classify Wine Bottles & Uncover Unique Words
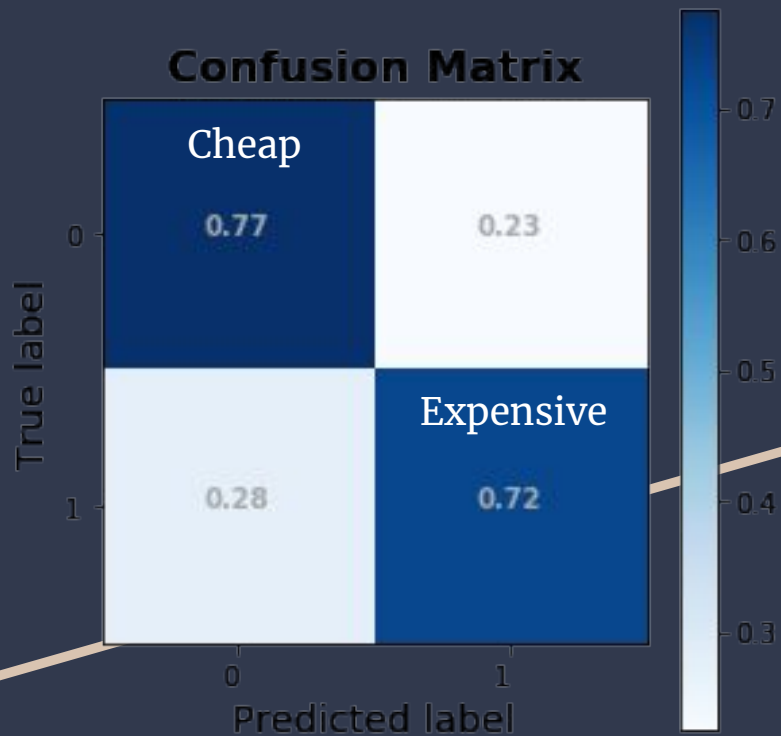
Acusio Bivona

Objectives

1.) Classify wine bottles by price
   - Done by examining wine reviews
   - 2 groups
      - Cheap vs. Expensive
   - Price cutoff - $25
      - Arbitrarily chosen
2.) Determine which words are most frequently used
   - Separated into 3 groups:
      - All of the bottles, the cheap bottles, and the expensive bottles
3.) Discover powerful business insights

Classification Results

Confusion Matrix

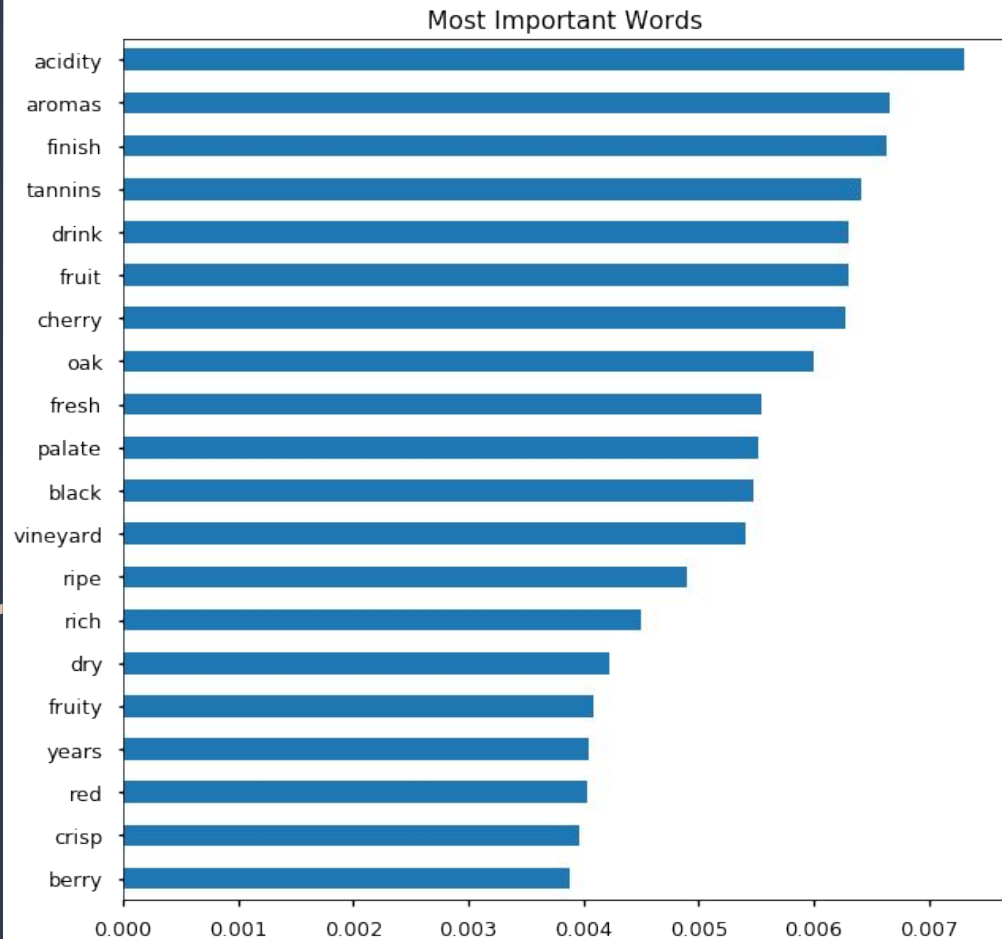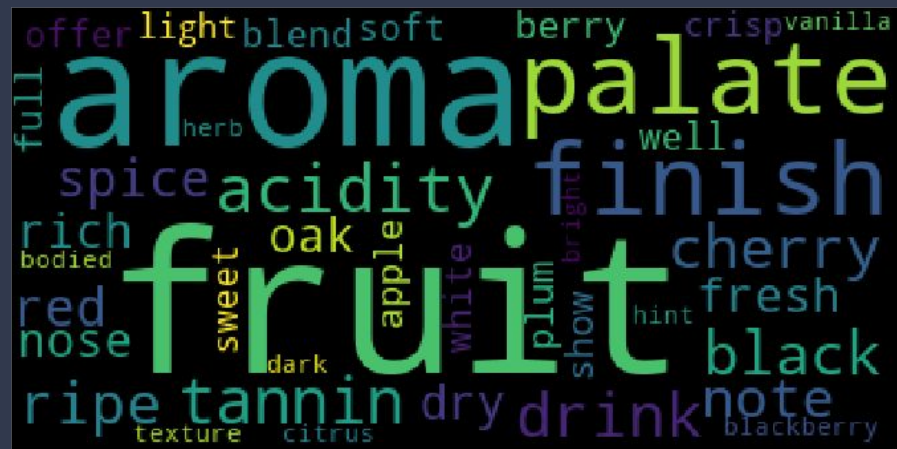|	Cheap (0)	Expensive (1)
True label 0	0.77	0.23
True label 1	0.28	0.72

Predicted label

- Top left = Cheap wines
  - Correctly classified with 77% accuracy
- Bottom right = Expensive wines
  - Correctly classified with 72%
- Overall Testing Accuracy: 75%
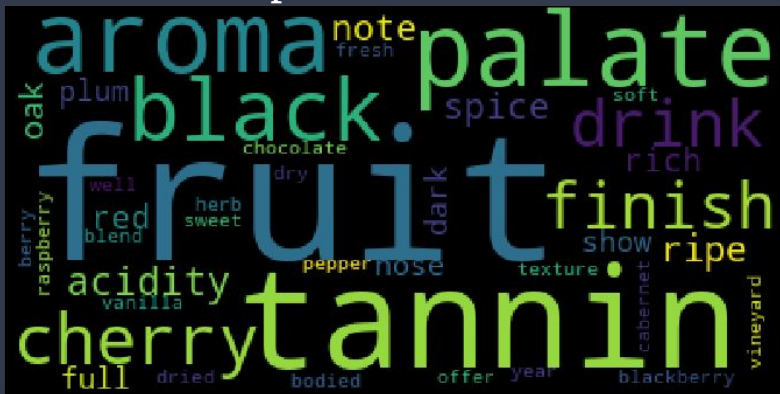 - Can perhaps increase performance with further experimentation

Important Words

Whole Dataset WordCloud (Frequency)
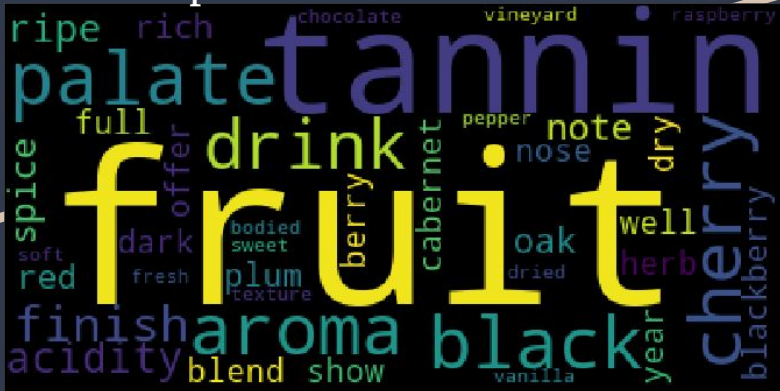

Most Important Words

Cheap vs. Expensive

Cheap Wine Wordcloud

Expensive Wine Wordcloud

Unique words for cheaper wines:

- Fresh
- Berry
- Crisp
- Dry
- Apple
- Fruits

Unique words for more expensive wines:

- Oak
- Rich
- Blackberry
- Cabernet
- Dark
- Blend

Significance: The two categories are distinct

Unique Word Pairings

Cheap Wine Reviews

Words	Frequency
('drink', 'now.')	0.00258455
('black', 'cherry')	0.00175851
('palate', 'offers')	0.00136917
('cabernet', 'sauvignon')	0.00103953
('sauvignon', 'blanc')	0.000949335
('ready', 'drink')	0.000945441
('green', 'apple')	0.000941548
('ready', 'drink.')	0.000891583
('stone', 'fruit')	0.000837076
('red', 'berry')	0.00082864
('berry', 'fruits')	0.000827342
('nose', 'palate')	0.000815662
('crisp', 'acidity')	0.000795546
('tropical', 'fruit')	0.000774133
('finish', 'drink')	0.000761155
('black', 'fruits')	0.000741039
('red', 'fruits')	0.000731306
('white', 'peach')	0.000726115
('black', 'currant')	0.000713137
('pinot', 'noir')	0.000712488

Expensive Wine Reviews

Words	Frequency
('black', 'cherry')	0.00299625
('cabernet', 'sauvignon')	0.00212652
('palate', 'offers')	0.00137533
('pinot', 'noir')	0.00125744
('black', 'pepper')	0.00120177
('drink', 'now.')	0.00113104
('cabernet', 'franc')	0.00107275
('petit', 'verdot')	0.000899201
('white', 'pepper')	0.000873004
('black', 'currant')	0.000850737
('black', 'fruit')	0.000840913
('french', 'oak')	0.000798999
('palate', 'delivers')	0.000798999
('red', 'cherry')	0.000784591
('firm', 'tannins')	0.000755119
('cherry', 'fruit')	0.00075119
('dark', 'chocolate')	0.000740056
('finish', 'drink')	0.000738091
('nose', 'palate')	0.000698796
('full', 'bodied')	0.000646403

Recommendations

- Able to obtain valuable insights on both marketing and new wine creation
    - Marketing:
        - Can highlight the unique aspects of the wine depending on its price point
    - Wine Creation:
        - I.e. Use qualities from expensive wine to increase value of cheap wine
        - Or, can use qualities of cheap wine to make expensive wines more alluring

Summary

- Confidently classified cheap vs. expensive wine using professional wine reviews
- Determined that cheap & expensive wines have unique characteristics
    - Cheap:
        - Berry, Dry
    - Expensive:
        - Blackberry, Rich

Future Works

If more time were permitted:

- Explore varieties of wine to discover their price points
- Explore wine regions
- Perhaps explore specific wineries

Thank you for your time!

Questions are welcomed